Economics Bulletin

Volume 31, Issue 3

The size distribution of all Cambodian establishments

Kiyoyasu Tanaka
Inter-disciplinary Studies Center, Institute of Developing Economies

Naomi Hatsukano
Area Studies Center, Institute of Developing Economies

Abstract

This paper describes the size distribution of all Cambodian establishments for 2009, showing that small- and large-scale establishments accounted for the largest share of employment. We find limited evidence for Zipf's law because Cambodian industry is characterized by a more dense mass of small establishments than the Zipf distribution would predict.

We thank Fumihiko Nishi and Net Kim for assistance on the Cambodian data, and the referee for improving the manuscript. We gratefully acknowledge the financial support from the Institute of Developing Economies.

Citation: Kiyoyasu Tanaka and Naomi Hatsukano, (2011) "The size distribution of all Cambodian establishments", *Economics Bulletin*, Vol. 31 no.3 pp. 2128-2137.

Submitted: Dec 20 2010. **Published:** July 20, 2011.

1. Introduction

The size distribution of business firms has received considerable attention for its striking empirical regularity. Axtell (2001) found that the size distribution of tax-paying firms in the U.S. is well described by a Pareto distribution with a shape parameter of 1. In other words, the probability that a firm has more than L workers is proportional to 1/L. A similar pattern was also found for a large sample of European firms by Fujiwara et al. (2004). Consequently, the regularity has yielded various theoretical explanations in which the underlying model of firm dynamics could generate an observed pattern of firm size distributions (Simon and Bonini, 1958; Luttmer, 2007; Rossi-Hansberg and Wright, 2007).

The firm size distribution in industrial countries has been widely examined, but there is the limited systematic analysis on firms in developing countries. Using aggregate measures of employment distribution by firm size, Tybout (2000) describes that a distinctive feature of manufacturing firms in developing countries is a bimodal structure in which a number of small firms and a handful of large firms account for a substantial share of employment. The observed feature is commonly attributed to high transaction costs due to imperfections in product and factor markets in developing countries (Sleuwaegen and Goedhuys, 2002). However, prior findings relied primarily on a limited sample of firms in a specific sector, so that there is limited understanding on the complete distribution of firms and the empirical validity of Zipf's law in the case of developing countries.

This paper extends the literature by presenting new empirical evidence on the entire size distribution of business establishments in least developing countries. Specifically, we use the first comprehensive *Establishment Listing* in Cambodia for 2009, which surveyed the economic activities of virtually all establishments in all areas of Cambodia. The analysis improves our understanding of business establishments in developing countries.

This paper is organized as follows. Section 2 describes the data and illustrates the size distribution of Cambodian establishments according to the fine disaggregation of employment size. Section 3 provides an empirical investigation of Zipf's law. Section 4 concludes.

2. Data

We use a unique dataset on Cambodian establishments from *Establishment Listing* in

Cambodia for 2009 (EL2009) surveyed by the National Institute of Statistics.[1] The survey defines an establishment as a unit of economic entity managed by a single ownership in a single physical location with some durable facilities. EL2009 covers all the establishments that were in operation in the entire territory of Cambodia as of February 2009, except for individual proprietorships in agriculture, forestry and fishery. The survey information includes location, employment, ownership type, and industrial category at the establishment level.[2] This dataset is novel in that it provides data on a fundamental unit of economic entity for the entire nonfarm private and public sectors.

Table 1 lists the size distribution of Cambodian establishments in the nonfarm sector by tabulating the number and share of establishments and employment over size classes. The sample has 375,854 establishments with 1,455,526 workers in Cambodia for 2009. Small-scale establishments with less than 10 workers account for over 90% of all establishments in number, with nearly a 60% share of nationwide employment. Medium-scale establishments between 10–99 workers constitute 3.2% by number and 16.8% by employment. By contrast, large-scale establishments with 100 workers or more represent only 0.18% by number, but account for a quarter of employment. As shown by Tybout (2000) for other developing countries, the size distribution of Cambodian establishments also exhibits a "missing middle" in which medium-scale establishments are underrepresented in the Cambodian economy.

3. Zipf's law

It is of interest to examine whether Zipf's law holds for the Cambodian establishments.[3] For a set of establishments $i = 1, \dots, N$, let S(i) denote the size of an establishment i, as measured by the total number of persons engaged in economic activity, including owners and/or family members. Assume that S(i) is a discrete random variable following a Pareto distribution. Then, the Pareto distribution of the size variable, S(i), is defined by:

$$f(S(i)|\alpha, s_0) = \frac{\alpha s_0^{\alpha}}{S(i)^{\alpha+1}}, \qquad S(i) \geq s_0, \alpha > 0 \tag{1}$$

$$F(S(i)|\alpha, s_0) = 1 - \left(\frac{s_0}{S(i)}\right)^{\alpha}, \quad S(i) \geq s_0, \alpha > 0 \tag{2}$$

where $f(\cdot)$ is a probability density function and $F(\cdot)$ is a cumulative density function.

[1] See details at http://www.nis.gov.kh/index.php/statistics/surveys/el2009.

[2] We exclude non-governmental organization (NGO) from the analysis.

[3] See Nitsch (2005) for empirical literature on Zipf's law for cities.

Table 1. Size Distribution of Cambodian Nonfarm Establishments for 2009

	Establishment		Employment	
Size	Number	Share (%)	Number	Share (%)
1	112,131	29.83	112,131	7.70
2	149,293	39.72	298,586	20.51
3	44,611	11.87	133,833	9.19
4	24,268	6.46	97,072	6.67
5	14,466	3.85	72,330	4.97
6	8,419	2.24	50,514	3.47
7	4,947	1.32	34,629	2.38
8	3,201	0.85	25,608	1.76
9	1,796	0.48	16,164	1.11
10-19	7,972	2.12	102,374	7.03
20-29	1,956	0.52	45,348	3.12
30-39	1,013	0.27	32,680	2.25
40-49	388	0.10	16,839	1.16
50-99	711	0.19	46,787	3.21
100 or more	682	0.18	370,631	25.46
Total	375,854	100	1,455,526	100

Notes: Size indicates the number of workers for each establishment; non-governmental organizations are excluded from the sample.
Source: Cambodian Establishment Listing 2009.

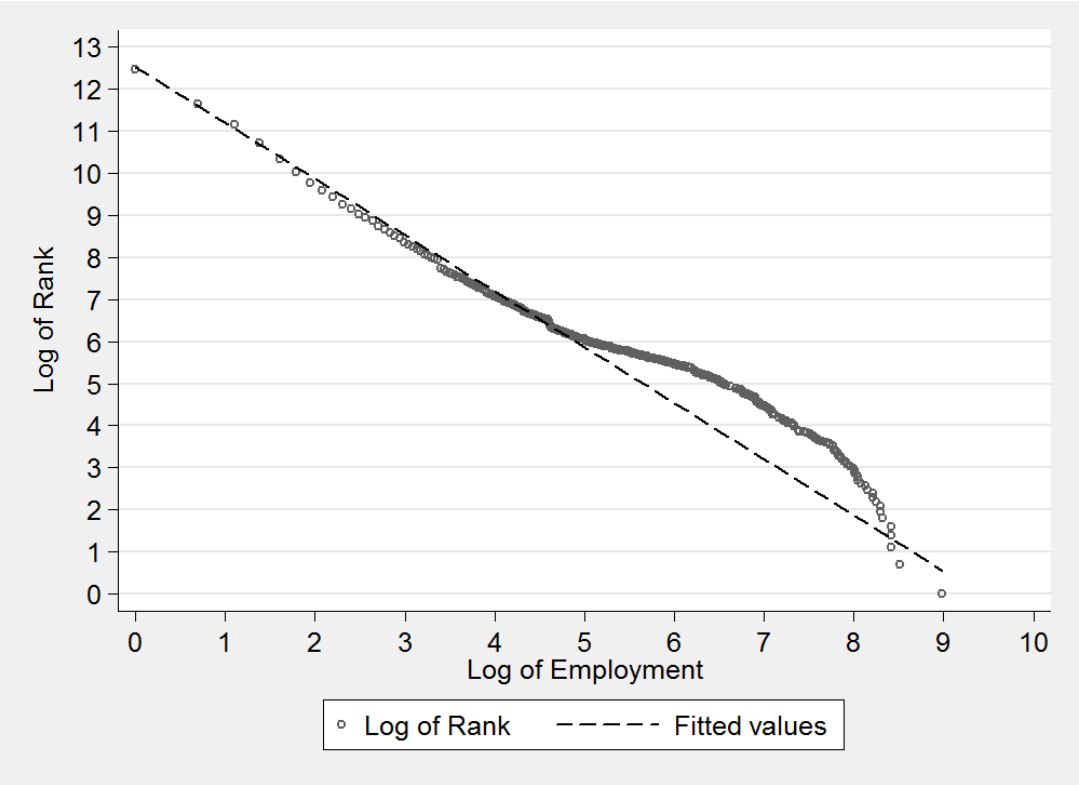
s_0 is the lower bound of the establishment size. α is a positive parameter that shapes the dispersion of the Pareto distribution. While the Pareto distribution describes the form of a power law in general, Zipf's law is a special case of the Pareto distribution with α = 1.

Zipf's law can be analyzed by looking at the log of the rank plotted against the log of the size. Let R(i) denote the rank of the size of establishments, S(i), which are sorted from largest to smallest. Because the rank is defined by $\mathrm{R(i)/N} = 1 - \mathrm{F(S(i))}$, it is expressed as:

$$\mathrm{R(i)} = \mathrm{N} \cdot \left(\frac{s_0}{\mathrm{S(i)}}\right)^{\alpha} \tag{3}$$

Taking natural logarithms, we obtain $\ln\mathrm{R(i)} = \beta - \alpha\ln\mathrm{S(i)}$, where $\beta = \ln\mathrm{N} + \alpha\ln s_0$ is a constant term. As explained in Gabaix (2009) and Gabaix and Ibragimov (2011), we try to reduce the small-sample bias by defining the dependent variable as $\ln(\mathrm{R_i} - 0.5)$.

Figure 1. Zipf Plot.



Notes: The circles indicate a scatter plot of Cambodian establishments; the dashed line is a linear regression line.

By allowing for an error term in the deterministic specification, the rank–size relationship is specified as:

$$\ln (R_i - 0.5) = \beta - \alpha \ln S_i + \varepsilon_i \qquad (4)$$

Based on the specification (4), we use an Ordinary Least Squares (OLS) to estimate α for the sample of Cambodian establishments. As the ranking procedure leads to a positive autocorrelation in residuals, a standard error for the estimate is calculated by $\left|\hat{\alpha}^{OLS}\right|(n/2)^{-1/2}$, where n is the sample size.[4]

Before proceeding with estimation results, Figure 1 shows a plot of lnR(i) against lnS(i) with a linear regression line. The plots in the range of 0–5 over log employment appear to fit well with the straight line, suggesting that the probability that an establishment has more than S employment is approximately proportional to 1/S. However, the plots in the range of over 5 log employment start to deviate substantially from the linear regression line. Thus, the graphical representation of the rank–size relationship does not seem to support the Zipf's law for the Cambodian data.

[4] The authors thank the referee for pointing out estimation issues.

Table 2. Regression Results of Zipf's Law

	Employment			OLS regression		
Sample	Mean	Std. Dev.	Obs.	$\hat{\alpha}$	Std. Err.	R^2
All	3.9	41.3	375884	-1.33*	0.003	0.994
Size:						
1-9	2.3	1.5	363159	-1.31*	0.003	0.994
10-99	20.3	14.4	12043	-1.29*	0.017	0.997
Over 100	543.4	800.7	682	-0.92	0.050	0.917
Sector:						
Manufacturing	6.2	83.8	84629	-1.30*	0.006	0.992
Wholesale and Retail	2.2	3.4	198103	-1.30*	0.004	0.994
Accommodation and food service	3.9	8.3	29225	-1.37*	0.011	0.995
Education	13.5	29.3	9020	-1.38*	0.021	0.994
Ownership:						
Individual proprietorship	2.8	21.2	358182	-1.32*	0.003	0.994
State-owned enterprise	13.9	33.0	8690	-1.37*	0.021	0.994
Formal company	74.1	346.0	2098	-1.25*	0.039	0.987
Foreign company	52.0	177.8	144	-1.24	0.146	0.983

Notes: The dependent variable is the log of the rank, and the independent variable is the log of the employment size; * indicates that the estimated coefficient is different from one at the 1% significance level.

We proceed to estimate the coefficient α. Table 2 presents the estimation results with summary statistics of each sample used for regression. Using the entire sample, an OLS estimate of α is 1.33, which is significantly greater than one at the 1% significance level. To extend the analysis, the sample is split along various dimensions. First, establishments are separated into three classes by employment size. The estimated coefficients are 1.31 and 1.29 for small-scale establishments (1–9) and medium-scale establishments (10–99), respectively. As these estimates are significantly different from one, the size distribution for small and medium establishments does not appear to fit Zipf's law well. By contrast, an estimate of α for large-scale establishments (100 workers or more) is much closer to the value of 1. This result will be further examined later.

Second, the sample is split by four major sectors in the Cambodian economy. The summary statistics show that average employment size per establishment for

manufacturing (6.2) is larger than that for wholesale and retail (2.2) and accommodation and food service (3.9). While the average employment size varies largely among these sectors, all the estimates of α are significantly different from one, ranging from 1.30 to 1.37. Finally, the sample is broken down by ownership type: individual proprietorship, state-owned enterprise, and private/public limited company. While the average employment size differs substantially by ownership, the estimated α is still significantly different from one in these samples; the insignificant result for foreign firms could be due to small sample size.

The results up to this point may be sensitive to the estimation method. To address this concern, we follow the approach by Axtell (2001) to estimate the probability density function (pdf) and cumulative density function (cdf) as specified in equations (1) and (2).[5] Following the literature, we divide the values of establishment employment into bins of equal size on the log scale. When 9 bins for $\ln S(i)$ are made with an interval of 1, the results show that the pdf estimate is -1.19 with the standard error of 0.07, and the cdf estimate is -1.19 with 0.05.[6] Although the estimates of α are relatively closer to 1, they are significantly different from one. Thus, the different methods do not provide strong evidence in favor of the Zipf's law.

For a final check, we examine the role of deviations for very small establishments from Zipf's law, as discussed in Gabaix (2009). Table 3 presents the estimation results for the sample with various minimum size cutoffs. For the sample with over 5 workers, the estimated coefficient is -1.29, which is significantly different from one. As we increase the threshold size over 25, the estimate becomes -0.99, suggesting that the Zipf's law describes the data well. However, the estimate deviates progressively from one in absolute value as we progressively increase the threshold size. While the Zipf law fits well the Cambodian establishment data above the certain minimum size, the fit is quite sensitive to the size threshold. This sensitivity is in stark contrast to the evidence in Axtell (2001) that the Zipf's law holds well for the size distribution of all U.S. firms.

In addition, the employment share is less than half of total employment when the sample fits the Zipf's law well. The exclusion of very small firms may be valid for firm size distribution in developed countries because their economic activity is relatively small (di Giovanni et al., 2010). However, very small establishments accounted for the substantial share of employment in Cambodia, suggesting that the sample is likely to be

[5] Giovanni and Levchenko (2009) use these approaches to examine large firm-level dataset from many countries.

[6] The log of employment size ranges from 0 to 8.98 with a standard deviation of 0.70.

Table 3. Sensitivity to the Minimum Size Threshold

	OLS regression				
Sample	α	Std. Err.	R^2	Obs.	Share (%)
Size:					
Over 5	-1.29	0.009	0.985	45564	55.9
Over 10	-1.14	0.014	0.984	12725	42.2
Over 20	-1.03	0.021	0.978	4752	35.2
Over 25	-0.99	0.024	0.977	3498	33.4
Over 30	-0.96	0.026	0.978	2796	32.1
Over 40	-0.94	0.031	0.967	1781	29.8
Over 50	-0.93	0.035	0.959	1393	28.7

Note: Share indicates the percentage share of employment for the sample in total employment.

unrepresentative of the economy when a small size threshold is imposed. Thus, we conclude that the Zipf's law fits only the limited segments of the size distribution of Cambodian establishments.

The regression analysis suggests that the size distribution of Cambodian establishments is not likely to be approximated well by Zipf's law. Along different dimensions of the sample, the estimated coefficient α tends to be significantly greater than one in absolute value. Although Zipf's law appears to hold for the samples with a certain size threshold, the sample size is unrepresentative of Cambodian establishments. These results imply that the size distribution of all Cambodian establishments is associated with a larger number of small establishments and a smaller number of large establishments than the distribution predicted by a Pareto distribution with a shape parameter of 1. This interpretation is consistent with the finding that the estimates of α tend to be lower for the sample with larger average employment sizes.

The large mass of small establishments in the size distribution could be a manifestation of substantial barriers to the growth of small- and medium-scale enterprises in the Cambodian economy (Sleuwaegen and Goedhuys, 2002). While monotonically increasing numbers of progressively smaller firms are also observed in developed countries as shown in Axtell (2001), an underrepresented share of mid-sized establishments in employment is a distinctive feature of the Cambodian industry. Possible deterrent effects on the growth of small establishments would range from regulatory barriers (taxes, registration fees, and corruption) to financial constraints on external credit, demand constraints on mass production, and infrastructure obstacles to transportation. An investigation of these determinants is beyond the scope of this paper.

4. Conclusion

This paper employs the first comprehensive data on Cambodian establishments to characterize the size distribution of establishments and investigate the empirical validity of Zipf's law in the context of least developing countries. The descriptive analysis shows that small-scale and large-scale establishments account for a majority of the number and employment of establishments in the Cambodian economy. In contrast, mid-sized establishments are underrepresented in the domestic industry, consistent with the "missing middle" that is commonly observed in a wide range of developing economies (Tybout, 2000). Additionally, the regression analysis provides considerable evidence against the strict validity of Zipf's law. Instead, the estimated shape parameter is generally greater than one, indicating that the dispersion of establishment sizes is relatively small with a more dense mass of small establishments.

References

Axtell, R. L. (2001) "Zipf distribution of U.S. firm sizes." *Science* 293, 1818–1820.

Fujiwara, Y., Guilmi, C. D., Aoyama, H., Gallegati, M., Souma, W. (2004) "Do Pareto-Zipf and Gibrat laws hold true? An analysis with European Firms." *Physica A: Statistical Mechanics and its Applications* 335, 197-216.

Gabaix, X. (2009) "Power laws in Economics and Finance." *Annual Review of Economics* 1, 255-293.

Gabaix, X., Ibragimov, R. (2011) "Rank – 1/2: a simple way to improve the OLS estimation of trail exponents." *Journal of Business and Economic Statistics* 29, 24-30

di Giovanni, J., Levchenko, A. (2009) "Firm, entry, trade, and Welfare in Zipf's world." Discussion paper No. 591, University of Michigan.

di Giovanni, J., Levchenko, A., Rancière, R. (2010) "Power laws in firm size and openness to trade: measurement and implications." IMF Working Paper.

Luttmer, E. G. J. (2007) "Selection, growth, and the size distribution of firms." *Quarterly Journal of Economics* 122, 1103-1144.

Nitsch, V. (2005) "Zipf zipped." *Journal of Urban Economics* 57, 86-100.

Rossi-Hansberg, E., Wright, M. L. J. (2007) "Establishment size dynamics in the aggregate economy." *American Economic Review* 97, 1639-1666.

Sleuwaegen, L., Goedhuys, M. (2002) "Growth of firms in developing countries: evidence from Côte d'Ivoire." *Journal of Development Economics* 68, 117–135.

Simon, H. A., Bonini, C. P. (1958) "The size distribution of business firms." *American Economic Review* 48, 607–617.

Tybout, J.R. (2000) "Manufacturing firms in developing countries: how well do they do and why?" *Journal of Economic Literature* 38, 11– 44.